İSTANBUL TECHNICAL UNIVERSITY ★ INSTITUTE OF SCIENCE AND TECHNOLOGY

FACE RECOGNITION USING HARMONIC IMAGES

M.Sc. Thesis by
Kıvılcım HELHEL

Department : Computer Engineering

Programme : Computer Engineering

OCTOBER 2009

İSTANBUL TECHNICAL UNIVERSITY ★ INSTITUTE OF SCIENCE AND TECHNOLOGY

FACE RECOGNITION USING HARMONIC IMAGES

M.Sc. Thesis by
Kıvılcım HELHEL
(504061519)

Date of submission : **07 September 2009**
Date of defence examination: **06 October 2009**

Supervisor (Chairman) : **Prof. Dr.Muhittin GÖKMEN(ITU)**
Members of the Examining Committee : **Prof. Dr. A.Coşkun SÖNMEZ (YTU)**
Assis. Prof. Dr. M. Ersel KAMAŞAK (ITU)

OCTOBER 2009

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

HARMONİK GÖRÜNTÜLERİ KULLANARAK YÜZ TANIMA

YÜKSEK LİSANS TEZİ

Kıvılcım HELHEL
(504061519)

Tezin Enstitüye Verildiği Tarih : 07 Eylül 2009
Tezin Savunulduğu Tarih : 06 Ekim 2009

Tez Danışmanı : Prof. Dr. Muhittin GÖKMEN (İTÜ)
Diğer Jüri Üyeleri : Prof. Dr. A.Coşkun SÖNMEZ (YTÜ)
Yrd. Doç. Dr. M. Ersel KAMAŞAK(İTÜ)

EKİM 2009

FOREWORD

I would like to thank to my supervisor, Prof. Muhittin Gökmen for his guidance throughout my research process and during the preperation of this thesis.

Special thanks to my family, for their patience and continuous support during my master study and during the preparation of this work.

I would also like to thank to TÜBİTAK (The Scientific and Technological Research Council of Turkey) for supporting me during my master study under the Grant "National Scholarship Programme for Master Science Students".

And finally, I would like to thank Birkan Tunç for valuable discussions and for sharing his knowledge with me.

October 2009

Kıvılcım HELHEL

Computer Engineering

TABLE OF CONTENTS

Page

ABBREVIATIONS

PCA	: Principal Component Analysis
LCA	: Linear Component Analysis
2DPCA	: Two dimensional PCA
LFA	: Local Feature Analysis
LDA	: Linear Discriminant Analysis
EBGM	: Elastic Bunch Graph Matching
HMM	: Hidden Markov Model
SVM	: Support Vector Machine
PPLS	: Parametric Piecewise Linear Subspace
LPCMAP	: Linear PCMAP
KCCA	: Kernel Canonical Correlation analysis
GDA	: Generalized Discriminant Analysis
LBP	: Local Binary Patterns
SHBMM	: Spherical Harmonic Based Morphable Model
ICA	: Independent Component Analysis
MAP	: Maximum a Posteriori Estimate

LIST OF TABLES

Page

LIST OF FIGURES

Page

FACE RECOGNITION USING HARMONIC IMAGES

SUMMARY

Face recognition is a task that people do consistently without any effort. The human brain causes the visual information of human face as a biometric identifier, because its apperance is highly informative and discriminative.

Face recognition using biometric information is a challenging function of the brain. Developing algorithms that try to act like brain does in face recognition is the one of the most important aim of the computer vision, pattern recognition and image analysis fields.

The face recognition problem can be explained as a given input image or video sequence, identify or verify one or more individuals in the input using a database that contains face images of known individuals. Because of development in technology and great increase in requirements of applications that use face recognition, the number of studies on face recognition increases rapidy in recent years.

Harmonic images based face recognition approaches are important in literature, because these applications greately facilitate the modelling of generic illumination. They are provide solutions to the face recognition problem under varying lighting conditions, especially outdoor environments.

In this thesis, a brief review of harmonic images (spherical harmonic images) and also an approach that uses the combination of statistical model and harmonic images are presented. This approach uses only one training image per subject. This is an important property for face recognition applications because most of the approaches dealing with illumination variation use more than one image. This approach and a similar approach that have some different parts from this- are investigated and experimental results of second one are presented. The experiments have been done using a database that includes synthetic images illuminated such as in YaleB database. According to the lighting direction, we seperated the images into four subsets that we use in both testing and training. Subset4 has the extreme lighting. For subset1 and subset2 the recognition rate is above 95%, for subset3, it is between 90% and 95% and for subset4 it is between 80% and 88%.

HARMONİK GÖRÜNTÜLERİ KULLANARAK YÜZ TANIMA

ÖZET

Yüz tanıma insanların hiçbir çaba sarfetmeden gerçekleştirdikleri bir görevdir. İnsan beyni yüze ait görsel bilgiyi, biyometrik tanımlayıcı olarak kullanır. Çünkü, insan yüzü oldukça bilgi verici ve ayırt edici bir özelliktir.

Biyometrik bilgileri kullanarak yüz tanıma, beynin zorlayıcı bir fonksiyonudur. Yüz tanıma konusunda beyin gibi davranmayı deneyen algoritmalar geliştirme bilgisayarlı görü, örüntü tanıma ve görüntü analizinin en önemli amaçlarından biridir.

Yüz tanıma problemi, girdi olarak verilen bir yüz görüntüsü ya da yüz görüntülerini içeren bir video kaydından, bilinen kişilerin yüz görüntülerini içeren bir veritabanını kullanarak, bir veya daha fazla bireyi tanımlamak veya doğrulamak olarak tanımlanabilir. Teknolojideki gelişmeler ve yüz tanımayı kullanan uygulamalara duyulan ihtiyacın artması nedeniyle, bu alanda yapılan araştırmalar son yıllarda hızla artmıştır.

Harmonik görüntü (küresel harmonik görüntü) tabanlı yüz tanıma yaklaşımları literatürde çok önemlidir. Çünkü bu yaklaşımlar genel aydınlanmaları modellemeyi oldukça kolaylaştırmaktadır. Değişik ışıklandırma durumları altında özellikle dış ortamda, yüz tanıma probleminin çözülmesini sağlamaktadır.

Bu tezde, kısaca harmonik görüntüler (küresel harmonik görüntüler), ayrıca istatiksel modeller ve harmonik görüntülerin bileşimini kullanan bir yaklaşım tanımlanmıştır. Bu yaklaşım, her bir kişi için sadece bir eğitim görüntüsü kullanır ve bu önemli bir özelliktir, aydınlanma değişimleri ile mücadele eden birçok yaklaşım bu amaçla birden fazla görüntü kullanmaktadır. Bu yaklaşım ve buna çok benzeyen, farklı bölümler içeren bir yaklaşım incelenmiş, araştırılmış ve ikincisine ait deneysel sonuçlar açıklanmıştır. Deneyler YaleB veritabanına benzer şekilde aydınlatılmış sentetik görüntüler içeren bir veritabanı kullanılarak gerçekleştirilmiştir. Test sırasında kullanılan görüntüler aydınlanma yönüne göre alt gruplara ayrılmıştır. Altgrup4 en aşırı aydınlanmayı içeren kısımdır. Altgrup 1 ve altgrup2 için tanıma oranı %95'in üzerinde, altgrup3 için %90-%95 arasında ve altgrup4 için %80 ile %85 arasındadır.

1. INTRODUCTION

Biometrics attract recognition technology because they can be used in any application which needs security and control access. In addition they eliminate risks with less advanced technologies, based on what a person has or knows, not who a person is. Biometrics means behavioural characteristics. Fingerprints and iris are the most common biometrics. There are many other behavioural characteristics that become to use in recognition, such as voice, signature and face which have been studied in a few decades.

Face recognition becames one of the most important biometric technologies. Because both iris recognition and fingerprint recognition may not be appropriate for people who do not want to colloborate. In addition iris recognition is expensive to implement. On the other hand, face recognition has many benefits over other biometric technologies. Its reliability and acceptance has a good harmony and it provides security and privacy. It can be performed easily in daily life. The nature of the face structure is very informative and discriminative. So, it is easily usable in recognition without bothering people. It has received the attention of the researchers from computer vision, image processing, computational neuroscience and pattern recognition because it is a highly successful application of these fields. The reasons of the popularity of face recognition applications in last decades are avilability of possible technologies and wide range of government, commercial and law enforced applications [1]. Face recognition applications are categorized as FaceID, access control, security, surveillance, smart cards, law enforcement, face databases, multimedia management, human computer interaction in [1]. Table 1.1 shows these categories and the scenarios of the example applications [1].

Table 1.1: Face recognition application categories and examples

Category	Example application scenarios
Access Control	Facility access, ATMs, computer access, computer program access, computer network access, online examination access
Security	Terrorist alert, secure flight boarding systems, stadium audiance scanning, computer and computer application security, internet security
Surveillance	Advanced video survelliance, nuclear plant surveillance, portal control
Smart cards	Stored value security, user authentication
Law Enforcement	Crime stopping and susspect alert, suspect tracking and investigation, criminal face retrieval and recognition
Multimedia Management	Face-based search, face-based video segmentation and summarization, event detection
Human Computer Interaction	Interactive gaming, proactive computing
Face ID	Driver licenses, immigration, national id, passport, voter registration

1.1 Face Recognition Systems

Face recognition problem is recognizing one or more people in a given input image or video sequence. It uses a database that includes face images of known people. Recognition from input image and recognition from video sequence may vary according to their recognition criteria, algorithms of segmentation and quality of the input(s).

Face recognition systems generally divided into three types.

1. Verification
2. Identification
3. Watch-list

Verification is a one to one match that compares the face image whose identity is claimed with template face images. The second type, identification, compares an input face image with all image templates in a face database to find out the identity of the input face. Lastly, watch-list compares an input face image with all the face images in the database, too, but it also computes a score for each comparison. Then the scores are sorted numerically from highest score to the smallest one. An alarm is raised when a similarity score is higher than a given threshold. To sum up, watch-list matches an input face to a list of suspects one-to- few matches.

A system that solves the face recognition problem usually consists of four parts.

1. Face detection
2. Face alignment
3. Feature extraction
4. Feature recognition

The face areas are segmented from background by using face detection. Face detection also provides close estimates of location and scale of each detected face. Face alignment is used for a more correct localization and normalized faces. Face alignment locates facial components such as mouth, eyes and nose. It uses geometrical transforms and normalizes the face image about properties such as size and pose. After the face image normalized geometrically, face extraction is applied to obtain effective information. The information is used for defining the differentiation faces of different people. This information includes important features of face image and named as feature vector. Face recognition is the verification and identification in the final part. The feature vector of the query face is matched to faces in the database so that the identity of the face is output if there is a match with enough confidence. If there is no match, the face is indicated as unknown face. Figure 1.1 shows the sketch of these steps [1].

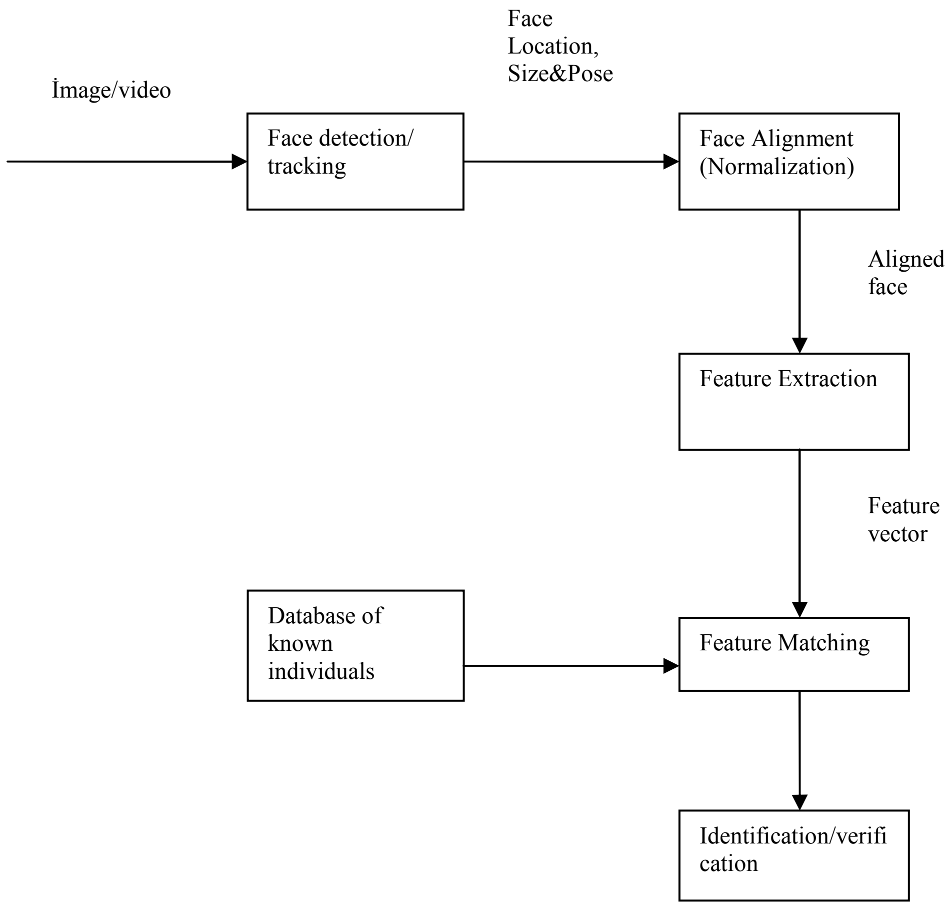


Figure 1.1: Block diagram of a generic face recognition system

1.2 Face Recognition in Subspaces

For face recognition, subspace analysis techniques are useful because the face images reside in a subset of input image space. Suppose that, an image of 80*80 has 6400 pixels that includes patterns. These patterns can be presented with 256^{6400} possible combinations of these pixels. Face can be one of the patterns in the input image and reside in a subset of the image space called as face subspace.

The most popular subspace methods are Eigenfaces and Fisherfaces. Eigenfaces is based on principal component analysis (PCA) and Fisherfaces is based on linear discriminant analysis(LDA). In this thesis, another subspace method that is based on QR decomposition is used.

1.3 Challenges in Face Recognition

There are so many sytems that are able to succeed in face recognition and many of them are able to achieve face recognition rates better than 80% [1]. Despite, there are still some points that can cause problems in face recognition processes. Because of recognition and examination process to find which class that face image is belong to, face recognition is a complex recognition problem. If images are not acquired under controlled circumstances, some problems that affect the recognition performance occur especially in systems that use only one image for training. The key problems are the sensibility of the classifiers to variations of circumstances. The most important ones are illumination and pose variations. There has been a huge effort to deal with these problems. Facial expressions, occlusions and the age variations are other problems that affect face recognition algorithms negatively. The number of studies that have been done in these problems are less with respect to researches on illumination and pose variations. The principle of dealing with illumination and pose variations is to develop robust classifiers.

1.3.1 Illumination variation

Lighting is not the same during a day or in every place. It changes time to time and from place to place. Because of face has a 3D structure, a direct lighting source causes strong shadows on face which reduce facial features, the part of the face that is exposed to light direction become highlighted and the other part of the face is reduced under the shadow such as in Figure 1.2. Illumination causes a wide variation even between the face images of the same class as explained as following "the variations between the images of the same face due to illumination and viewing direction are almost always larger than the image variation due to change in face identity" [2].

(a) (b) (c)

Figure 1.2: Same face under different lighting conditions. In (a) the face is illuminated by frontal light. In (b) the face is illuminated from left side. In (c) the face is illuminated from right-side.

To eliminate the effects of ordinary illumination variation in face recognition fundemantally a bidirectional method is applied. In this method, after applying a normalization process to input images, a robust face extraction method is developed to deal with illumination variation. Mean value normalization, histogram equalization and illumination correction are the most popular normalization methods[1]. Linear stretching is an example of an intensity normalization operation.

In linear stretching, a linear transformation is used and the original range of the query image is changed into a specified full image. Histogram equalization increases the contrast of the input image without affecting the global contrast. Using this way, the strength of the resulting image is distributed on the histogram and this decreases the strong illumination. Illumination correction is a least-squares method, strong illumination is decreased in the difference image. In this method firstly, I'(x,y) the best fitting plane is obtained as follows,

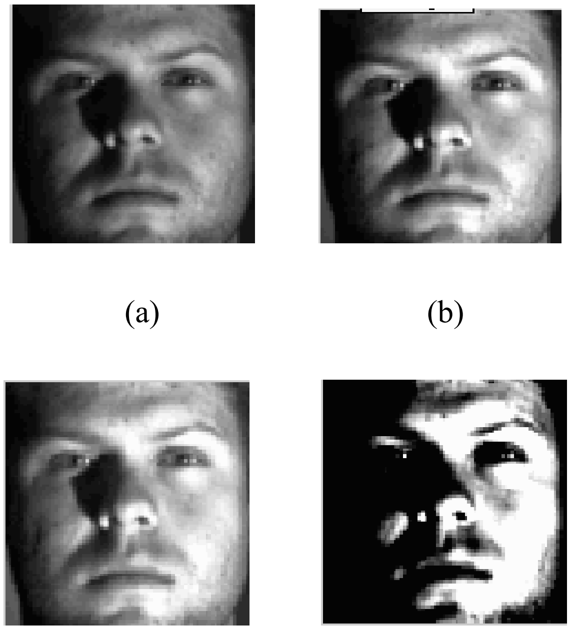
$$I'(x,y) = a \times x + b \times y + c, \tag{1.1}$$

where the values of a, b, c can be estimated using a least-squares method, than the illumination is corrected in the resulting difference image as follows,

$$I''(x,y) = I(x,y) - I'(x,y), \tag{1.2}$$

In Figure 1.3 the effects of histogram equalization, linear stretching and mean-value normalization methods can be seen.

Dealing with illumination variation is a control topic in computer vision. Therefore, there are lots of suggestions for illumination invariant face recognition. The main aim of this thesis is to eliminate the effect of the illumination variations in face recognition. So, the studies that try to deal with illumination variations will be explained in more detail in Section 1.3.



(a) (b)

(c) (d)

Figure 1.3: Normalization examples (a) Original image from YaleB (b) Linearly stretched image (c) Histogram equalized image (d) Zero-mean normalized image

1.3.2 Pose variation

The pose of the testing and training images are usually not the same in face recognition sytems. For example, the training image may be frontal while the testing image may be taken from a camera placed in the room, viewing the person from an angular position. Pose changes cause projective deformations and self-occlusions. Therefore, they affect the identification process negatively and challenged the researchers working on the face recognition systems. In face recognition under arbitrary pose variations, the problem of developing algorithms to recognize a face in a different viewpoint is being investigated. There are some methods which deal with up to a certain head rotation. However in some cases such as when security cameras create viewing angles that are out of this range, these methods do not deal with this problem.

Linear subspaces have an important place in face recognition under arbitrary pose variations and lots of works have been done that extend linear subspaces to solve this problem. A method was presented in [3] for face recognition with 3D pose variations. It was developed by Okada et al. . In this method, each known people in the training set is represented by using parametric linear subspace model. They stored the parametric linear subspace model representations of each people in which, each model can be fit to input resulting faces of known individuals whose head pose is aligned toinput face. They have worked on two different linear models: the linear PCMAP (LPCMAP) and the Parametric Piecewise Linear Subspace (PPLS) models. The LPCMAP is a parametric linear subspace model that combines the linear subspaces that are spanned by matrices where matrices connect projections coefficients of training samples on the subspaces and their corresponding 3D head angles. The PPLS is an extended version of the LPCMAP by means of the piecewise linear subspaces, which is a set of linear models. Each model supplies continuous analysis and synthesis mapping and makes it possible to generalize about unknown poses by interpolation. It is showed that the recognition system is strong against large 3D head pose variations. The PPLS system performs better than the LPCMAP because it makes the data size fill much less space.

The light-field, that is a 5D function of position and orientation, is used by Gross et al. in [4], in order to achieve a better robust and stable face recognition in dealing with the pose variation problem. The light-field means the radiance of light in free space. In this method, a set of eigen light-fields is obtained and the PCA is applied to a collection of light-fields of faces of different subjects. However, the average light field could also be estimated and deducted from all the light fields. After eigen coefficients can be estimated, input face images are vectorized in light-field vectors. Then, the vectorized face images are used for training and testing the system, which performs the standard eigenfaces algorithm.

1.3.3 The occlusion

Besides illumination and pose variations in face images, occlusion in the face images causes an important change in the face images. The occlusion on the upper-side of a face decreases the performance of the face recognition more than the occlusion on the lower-side of the face. Local approaches that divide a face into different parts and use a voting space to find match are the most popular ways to solve the problem of partially occluded objects.

The voting tecnique may make mistakes in classification, because it does not understand how good a local match is. To solve this problem, in [5] Martinez et al. has been introduced a method. The method divides each face image into k different local parts, and models each of face by using a Gaussian distribution. Gaussian distribution is for explaining the localization error problem. The mean feature vector and the covariance matrix for each local subspace are drown out and the probability of a given match can be connected with the sum of all distances. This approach is successful in only occluded faces.

Kurita et al. developed a method using neural networks that rebuilds the occluded part and discovers the occluded areas in the input image and this method was presented in [6]. The network is trained on non-occluded images in normal conditions, and during the testing, the occluded areas can be replaced with the recalled pixels and the original face can be rebuilt. They showed that classification performance does not change even when 20-30% of the face images is occluded.

In [7], Sahbi and Boujemaa has presented a complete face recognition algorithm based on feature extraction in difficult conditions. Firstly, the features are extracted, then they build a binary image that is subdivided into areas and describe a shape variation between faces. They model the statistical difference of each feature in the face model concerning its corresponding matched features in each candidate face of each training set. A matching class, which expresses the probable difference of this feature concerning the training images, is obtained. For small occlusions and rotation effects, the matching process do well and the recognition rate is high and unchangeable.

1.3.4 Time delay

Another problem in face recognition is the time delay, because the face changes over time in a non-linear way. If the time lapse is not very small between getting training and testing images, the performance of most recognition techniques is decreasing. Time delay problem is usually harder to solve than the others. It is still an unexplored and interesting aspect to improve the face recognition systems so as to make them deal with the changes in age. Only a few works have been done about this problem.

There are a few methods that periodically upgrade and retrain the system to deal with this problem. Because of the identification performs frequently, this is not an appropriate and practical solution. Another method is simulating the age of the subject to make system more robust about this type of variation.

In [8], Lanitis et al. have proposed a method that try to cope with this problem. The age functions that is obtained using a set of parameters that describe every image in the face database are used in this method. The best age function is choosen for each subject according to these parameters. Different subject-based age functions allow to consider about external factors which cause to the age variation and this is the most important advantage of this approach. The mean age of the subjects has been simulated in both the experiments and then the recognition task has been performed.

1.3.5 Face expression variation

Expression variations also causes an important amount of change in facial appearance. The developed algorithms are quite robust to facial expressions except from extreme expressions, for example screaming.

In [9], Donato investigated several methods for classifying twelve facial actions. He showed that Gabor filters based and Independent Component Analysis (ICA) methods performed the best among other methods such as Local Feature Analysis (LFA), LDA and Local PCA.

In [10], Tian et al. presented an Automatic Face Analysis system to analyze facial expressions based on both permanent features (eye-brows, eyes, mouth) and transient facial features in a nearly frontal-view face image sequence. They reported recognition rates of 96.4 percent for upper face action units and 96.7 percent for lower face units.

In [11], a study investigating the effects of facial asymmetry in face recognition under varying expression is presented by Liu et al.. They have showed that quantified facial asymmetry improved face recognition significantly when combined with conventional methods such as Fisherfaces and Eigenfaces

In [12], facial expression recognition by Kernel Canonical Correlation Analysis (KCCA) is proposed by Abate et al.. They manually locate 34 facial expressions in each image and using Gabor filters they transform these points into a labeled graph. Then, for each training image, they formed a six-dimensional semantic vector describing basic expressions. Learning the correlation between the semantic vector and labeled graph vector is achieved by KCCA. They presented better results than conventional approaches like LDA and GDA.

1.4 Illumination

Illumination changes influence the performance of face recognition methods negatively as explained shortly in Section 1.2.1. To solve the illumination problem a pretty large number of methods are investigated. The fundemantal methods can be grouped as: The shape from shading approaches, regenerative methods and representation methods. First and second derivative of the grey level images, 2D Gabor filters and edge maps are exemples of these aproaches respectively. All these techniques, however, failed to deal with the problem by themselves. Therefore some new methods are developed which extend these methods as follows.

In [13], with a new approach the edge map technique was improved by Gao et al., called Line Edge Map. In Line Edge Map face contours were determined and joint in segments. Then, these segments were organized in lines. The Houssdorf distance was changed a bit, to manage the new feature vectors. Besides, a new prefiltering technique was described for screening the whole set of subjects before performing the real testing operation. This method has been tried out in several different poses and illumination. Although this approach outperforms some other methods, for instance linear subspaces or Eigenfaces [14], the Fisherfaces [14] are stil superior as they can maximize the between-subject variability and minimize the within-subject differences.

In [15], Sim et al. presented a model and exampler-based approach for recognition. The main idea was to synthesize many examplers, that are used in the learning stage of a face recognition system. To realize this, they developed a statistical shape-from-shading model to recover the face shape from a single image, and to synthesize the same face under new illumination. The aim of using this is to build a fast classifier that was not possible before because of lack of training data. They used three classifiers: the first classifier performs nearest-neighbour search directly, the second classifier searches for the nearest exampler in the global PCA subspace and the third classifier computes individual PCA subspaces from the examplars. The accuracies of the first two classifires are not good enough and the accuracy of the third one is 95% by using YaleB and CMU-PIE databases.

In [16], Georghiodes et al. suggested an illumination cone model for face recognition under varying lighting. In the image space, under arbitrary lighting, the set of images of an object forms a convex cone and this fact was used in this model. It was also explained how the cone was made from as few as three images which are all under small lighting changes. The distances are calculated between the test image and each illumination cone, and the face which has the shortest distance is thought as the true identity.

In [17], Romdhani et al. recovered the shape and texture parameters of a 3D morphable model in an analysis by synthesis manner. This method compute a statistical model during a bootstrap phase which encapsulate texture and shape parameters.

An illumination invariant method which was developed from near-inframed (NIF) images is proposed by Li et al. in [18]. They used Local binary patterns (LBP) and took out features from the near-inframed images. NIF imaging contributed to the monotonic grey level transform to which LBP provided a good solution. Also, they used ada-boasted LBP features and improved a face matching engine, and presented and LDA-like scheme to further select discriminative LBP.

Harmonic images based face recognition methods are methods that are successful under arbitrary unknown illumination conditions and most of them uses only one training image per person. In Section 1.5 the algorithms that use spherical harmonic images for face recognition will be presented.

1.5 Harmonic Images

Harmonic images (spherical harmonic images) model is an important development on light modeling. Many studies have done after the studies of Basri et al. [19] and Ramamoorthi et al. [20].

The first study on turning incoming light into reflection was done by D'Zmura [21]. He showed that describing the reflection in terms of spherical harmonics.With this representation, after truncating high-order components, the reflection process can be written as a linear transformation, so the low-order components of the lighting can be recovered by inverting the transformation. He used this analysis to explore ambiguity in lighting [10].

The most important work in face recognition by using spherical harmonics is done by Basri & Jacobs [19] based on [21]. They showed that, nine or fewer basis images were enough to approximate the images of a convex Lambertion object with high accuracy. In addition, this method was successful to deal with the attached shadows. Basri & Jacobs used a kernel which represents Lambert's reflectance to model the reflectance functions in analogous with the convolution of each lighting function. They proved that the set of images of a convex Lambertian object could be approximated by a nine-dimensional subspace that include nine harmonic images. They explained that the first nine harmonics includes 99.2% of the energy of kernel and clarified how to get the nine-dimensional subspace from a model of an object which contains albedo and 3D structure. This method will be explained in more deail in Section 2.

The following methods that will be presented, have developed from the study of Basri & Jacobs in [19]. Zhang et al. focused on face recognition with spherical hormonics in [22]. The approach that belong to them will presented as follows: They suggested a method based on spherical harmonics that needs only one training image per subject. This method requires no 3D shape information, because it builds the statistical model based on a collection of 2D basis images. They showed that they can estimate the spherical harmonics basis images from just one image , by using the learned statistics, taken under arbitrary illumination conditions if the pose is invariant. They use YaleB database in experiments and the accuracy was changed between 97.2% and 100% with respect to illuminant angle. This method will be discussed later in more detail.

In[23], Zhang et al., represented two methods for face recognition under arbitrary lighting by using spherical harmonics and compared them. The first method is same in [22] and the second method constructs the statistical models directly in 3D spaces by using both the spherical harmonic illumination representation and a 3D morphable model of human faces. Using the statistical model the basis images are obtained. The recognition method in [19] is used for both two methods. With respect to previous works done in this field, recognition rate is good. But when the poses of training and testing images are very different, the recogniton rates are not as good as under constant pose.

In [24], Zhang et al. proposed a 3D Spherical Harmonic Basis Morphable Model (SHBMM). They obtained this model by integrating spherical harmonics into the morphable model framework and demonstrate that any face under arbitrary lighting can be simply represented by three low-dimensional vectors: shape parameters, spherical harmonic basis parameters and illumination coefficients. Given an image of a face, they explicitly recovered the shape information and estimate the spherical harmonic basis of the face. In this method, testing and training images can have different skin albedos and poses. In addition their model detects the shadow errors from the image difference between the testing image and the rendered image. They can remove the cast shadows from the input image and add cast shadows to the image to generate more realistic images. They experimented their method by using CMU-PIE and USF databases. They achieved high recognition rates for images under a wide range of illumination conditions.

In [25], Rara et al. represented a face recognition method using statistical model that combines shape (2D and height maps), appearence(albedo) and spherical harmonic projection information. The method takes a 2D frontal face image under arbitrary illumination as an input and outputs the estimated 3D shape and appearence. Face identification is performed using the shape and albedo coefficients. For frontal images the recognition results are perfect on Extended Yale B database.

In [26], Qing et al. presented an illumination normalization approach by relighting face images with a canonical illumination based on the harmonic images model. Using the harmonic image model in [23], they presented a model-based approach to identify faces robustly under generic illumination with the face relighting technology. Face recognition is achieved by matching the canonical form of the test image with canonical form of the training images. The experimental results that obtained using YaleB are little worse than that of the illumination cone [16] and harmonic examplers [22]. After relighting under arbitrary illumination direction, the error rates are decreased. As a result, relighting strategy improve the performance of the face recognition approach.

In [27], Yue et al. presented a method that extended the spherical harmonics to encode the pose information into the harmonic representation. They showed that the basis images of a rotated test image is a linear combination of the basis images at the frontal pose. They used the learning method that is the same in [22] to recover the harmonic basis images from only one image. For a rotated testing image under arbitrary illumination condition, a correspondance is applied between this testing image and the training images. Then the frontal pose image was warped from the testing image and the face was identified by using a linear construction based on basis images that is the closest to the testing image. They experienced the method on Vetter's 3D face model database. They used synthetic images. The recognition rate was nearly perfect, except extreme illumination conditions.

1.6 Face Recognition From a Single Intensity Image

In face recognition field, apperarence-based methods are the dominant ones since 1990's. Performance of these methods is depending on the number of training samples for each subject. Most of current face recognition methods use at least two samples of the same subject. On the other hand, in real-world applications, the number of samples per person is smaller than the assumed number of samples. In many applications such as driver licence one training sample per subject is used in the database. Therefore, the techniques that are using one training sample per person are studied in last decades. The advantages of one training sample per person in face recognition systems are collecting the samples easily, decreasing the store cost of samples and decreasing the computational cost of the recognition system [29]. Those are desired by the real-world applications.

Face recognition methods from intensity images have three categories: holistic methods, local methods and hybrid methods [29]. This algorithms are categorized depending on the sort of features used by different methods.

1.6.1 Holistic methods

Holistic methods use the whole image as input for identifying a face and concatenate the gray-values of all pixels in the face. In these methods, each face image is represented as a simple high-dimensional vector. Holistic methods keep all the detailed texture and shape information, which is useful to recognize the differences between faces. Under the condition of one sample per person, this representation causes two adverse results. A trade-off between high dimensions of image data and small samples more serious is appeared, when holistic methods are used. The most important problem about these methods is how to deal with very small sample problem. Also, the within-class variation cannot be calculated directly because there are only one vector for each class.

There are two ways to deal with these problems. One of these ways is to squeeze the information as much as possible from the simple face image. It can be done in the high-dimensional face space or in the dimensionality-reduced eigenspace, as extensions of the standard PCA technique. The other way is to build new views or different representations for each image and to include prior information. By this way, the real training set can be made much more bigger.

PCA and LCA are two examples of holistic methods. Much success has been fulfilled by holistic methods, however only one feature vector is used in representing each face image. This representation is sensitive to large appearance because of expression, illumination, pose and partial occlusion.

If only one sample per person is given, the generalization performance of PCA is not enough. But PCA is applicable for any sample size based on the viewpoint of computation. To get higher performance, the method have been extended by Wu et al. in [30]. They presented a method named $(PC)^2A$ to strenghten the information of face space. They obtained the projections of image that reflect the distirbution of the facial features. These features are more useful for face recognition than the features obtained from PCA. Then from the projected images, a new image is synthesized and it is combined with the real image. The aim of this act is to obtain a strengthen procedure. After this process, unimportant features are faded out and the important features become more salient.

A potential solution to reliably estimate the covariance matrix under the small sample size condition, two-dimensional PCA (2DPCA) has recently proposed by Yang et al. in [31]. This method uses 2D image matrices rather than 1D vectors for covariance matrix estimation. So, for small sample size problem, it is computationally cheap and suitable. The good results of 2DPCA has been observed on several well-known image databases.

In[32], Jung et al. develeoped a system to handle the expression, illumination and pose problems. They try to synthesize multiple new face images which imitate the corrupted images for recognition.The error rate of this method is 1.32%. This method improve the similarity between the corrupted images and the training images.

In [33], Penev et al. proposed Local Feature Analysis (LFA) approach. The aim of this approach is to cope with the limitations of PCA such as lack of providing local features and production of global non-topographic linear filters, in this method, a dense set of local feed forward receptive fields defined at each point of the receptor grid, whose outputs were as decorrelated as possible, were derived. The residual correlations contained by these outputs were further used in the sparsification of the output. The final representation was a local sparse-distributed representation in which only a small number of outputs are active for any given input.

1.6.2 Local methods

Local methods use the local facial features as a difference from holistic methods for recognition. Deciding how to include global information into local face models is a difficult subject in these methods.

Local methods are more suitable than holistic methods for dealing with the one sample problem. There are three reasons for this feature of local methods [29]:

1. The original face is represented by a set of low-dimensional local feature vectors, not by one simple full high-dimensional vector.
2. Local methods supply additional flexibility to notice a face based on its parts.
3. Different facial features can rise the variety of the classifiers and it is useful for face identification.

The local feature-based method and the local appearance-based methods are the categories of local methods [2]. Local feature-based methods determine local features firstly, and then extracts features on the located feature points. Usually only one simple image is used to obtain geometrical measures such as width of head or distance between eyes. These measurements are stored in the database to be used later. Most of earlier face recognition methods are belong to this category. Gabor wavelet decomposition and Elastic Graph Matching are examples of this group of methods. Local appearance-based methods simply divide the face image into sub-regions based on which local features are directly taken out. Local region partition, feature extraction, feature selection and classification are the steps of local appearance methods. Local regions are defined in local region partition step. In local feature extraction step, it has to be decided how to represent the information of them, which is important for the performance of a recognition system. Grey-value features and different types of derived features, such as Gobor wavelets are frequently used features. Additional feature selection stage is used for effectiveness and efficiency when a lot of features are created in the previous step. PCA and LDA are common feature selection methods and the most common way for identification is combining classifiers.

In[34], Brunelli et al. presented a method that uses a set of templates to detect the eye position in a new image by looking for the maximum absolute values of the normalized correlation coefficient of these templates at each point in the text image. There are many methods that have similar aim and operations.[35-37]

In [38], Manjunath et al. presented a method for feature detection and representation of face. The method is based on Gabor wavelet decomposition in [39]. Location information and feature information for each detected point is saved and model the relationship between the feature points a topological graph has been constructed. After that, the face recognition problem can be handled as graph matching problem [40] by Wiskott et al.. The recognition rates of this method is exceeds 86%. The disadvantage of this method is not to allow the modifications in using databases because it uses topology graph.

In [41], Lades et al. presented a deformable topology graph matching method, that is known as Elastic Bunch Graph Matching(EBGM) and become a successful method in the literature. Similar to [38], a topology graph is constructed based on Gabor filters. The Gabor features are robust against illumination change, distortion and scaling[39].

In[5], Martinez et al. presented a local probabilistic approach to recognize partially occluded and expression variant face from a single sample per face. The face subspaces are obtained using local patches at the same position of each face image, than they transformed into eigenspaces. In the identification part, testing image is divided into subregions and projected to eigenspaces. To measure the similarity of a given image to the training images, a probabilistic calculation is used. Under synthetically generated images the results shows that the local probabilistic approach does not reduce accuracy even when 1/6 of face is occluded.

In [42], Nefian et al. developed a Hidden Markov Model (HMM) based method. In their method, the observation vectors of the HMM system were extracted by Karhunen-Loeve transform. They presented an 86% recognition rate on ORL database. They also proposed a novel HMM-based face detection method in the paper.

1.6.3 Hybrid methods

Hybrid methods combines the former two methods. They use both holistic and local features, which makes the performance of a recognition system better than individual local or holistic methods. The most important factors affecting the performance of hybrid methods are how to determine which features will be combined and how to combine them in order to keep their advantages and prevent their disadvantages at the same time. Local features and the global features have different properties and can offer complementary information about the classification task. This method is not appropriate for recognition that uses only one training images. Therefore, hybrid methods might be a good way to reduce the complexity of classifiers and make their generalization capability greater.

In [8] Lanitis et al. proposed a flexible appearance based method for face recognition that used shape and intensity information. The statistical shape model is trained on 2D images using PCA. In classification, between-class variations of the shape model are differentiated from the within-class variations of the shape model by discriminant analysis. Local gray-level models were also built on the shape model to cope with the local appearance changes such as local occlusions. A global shape-free representation was also obtained by using mean shape and PCA. Finally, these three representations, shape parameters, shape-free parameters and local features were used together to compute a Mahalanobis distance.

In [43], Huang et al. proposed a component-based detection and recognition system. Component-based methods seperate face into several components such as mouth, eyes, nose that are connected by a flexible geometrical model. By using components, the changes of the head pose would affect the components positions and this would be coped with the flexibility of the geometric model. A disadvantage of the system is the need of a high number of training images containing different pose and illumination variations. In the classification phase, SVM classifier was used. On a set of six subjects; training on 3 images and testing on 200 images, the hybrid system presented a recognition rate of 90 percent.

In [12], Abate et al. represented that the local and global features are separately sensitive to different variation factors. For instance, local features may be affected more by illumination changes and global features may be influenced more by expression changes. In Table 1.2 properties of the features are summarized [12].

Table 1.2: Comparison of the sensitiveness of local and global features to variations

Variation factors	Local Features	Holistic Features
Small variations	Not sensitive	Sensitive
Large variations	Sensitive	Very sensitive
Illumination	Very sensitive	Sensitive
Expressions	Not sensitive	Sensitive
Pose	Sensitive	Very sensitive
Noise	Very sensitive	Sensitive
Occlusion	Not sensitive	Very sensitive

The studies on this part is still a few because the methods are not suitable for one sample problem [12].

1.7 Organization of Thesis

In Section 2, general information about the lambertian reflectance and the spherical harmonics is given. In Section 3, face recognition method that uses spherical harmonics and a single training image per person is explained in detail. In Section 4 the method that we worked on is totally presented. In Section 5, experimental results are presented and in Section 6 conclusion and feature work are given.

2. LAMBERTIAN REFLECTANCE AND MODELLING ILLUMINATION VARIATION WITH SPHERICAL HARMONICS

Spherical harmonics are used to cope with illumination variation by modeling the illumination. Most of algorithms that try to solve this problem, are not suffice for modeling illumination variation in uncontrolled conditions like outdoors. The obvious empirical method for analysing illumination variation is recording a number of images of an object under light sources from all different directions. This corresponds to moving a light source in a sphere around the object while the camera and the pose are fixed. Because the light moves in a linear way, the image under multiple light sources can be showed as the linear combination of source images that are obtained under individual light sources. So, instead of using images, the linear combinations or basis images that explain the illumination variation can be used. Usage of at least four basis images is enough for modeling the variability due to the illumination, generally nine basis images are used.

The works on face recognition methods that use spherical harmonic images gain a speed after the works of Basri and Jacobs [19] and Ramamoorthi and Hanrahan [20], because they showed that the set of images of a convex Lambertian object which is obtained under arbitrary illumination conditions can be approximated by a nine-dimensional or fewer subspace. The subspace includes nine basis images or fewer images of the object called as harmonic images. The method in [19] is appropriate for dealing with the attached shadows because the attached shadows can be accounted for with analytically derived linear subspaces.

In Section 2.1, a background information is given about reflectance and lighting, in Section 2.2 the previous works done for attached shadows are discussed, then the face recognition method based on spherical harmonics is explained step by step that developed by Basri et al. [19].

2.1 Reflectance and Lighting

There are common assumptions in handling illumination in computer vision and computer graphics. The first one is assumption of distant illumination to represent lighting as a function of direction. The second one is to neglect cast shadows from one part of the object on another. In the following sections, these assumptions are taken into account.

While explaining reflectance and lighting we assumed only distant light that shines on each point in the scene that have the same intensity (the surface of the object reflects the light according to Lambert's law). According to Lambert's law [24] objects reflect the light uniformly.

For the illumination from a point light that has a single direction on a Lambertian object, a model is considered. In this model, illumination can be represented by the unit vector u_l and intensity l. If there are multiple light sources, the intensity of light can be described as a funcion of direction $l(u_l)$. In this case, the intensity of light does not depend on the position in the scene and it can be assumed as a non-negative function of the surface of a sphere.

Building accurate face models from reflecting light is a complex task. When the light entering the skin at one point, it may exit another point, so the skin is not homogenous. This situation is not an easily solvable problem. So, the simpler models are more applicable and efficient for face recognition by considering the assumptions [1].

In this model the only variable is albedo of the object. The albedo describes the fraction of the light reflected at that point. According to Lambert' law and the given information up to now, the intensity (radiosity, reflected light) is obtained as follows:

$$i = l(u_l)\rho \max(u_l.v_r, 0), \qquad \textbf{(2.1)}$$

where l corresponds to a light ray of intensity, u_l corersponds to direction of the light, ρ is the albedo of a surface point and v_r is the normal direction.

By ignoring albedo and assuming the light direction is constant, the reflected light is a function of surface normal only. If light reaches a point from multiple directions, the light reflected by the point is obtained by taking integral over the contribution for each direction.

$$r(v_r) = \int_{S^2} k(u_l.v_r) l(u_l) du_l \, , \qquad \textbf{(2.2)}$$

where $k(u_l.v_r) = \max(u_l.v_r, 0)$ and $\int_{S^2}$ indicates the integration of the surface of a sphere.

2.2 Spherical Harmonic Representation and Properties of Convolution Kernel

Until Basri et al. and parallely Ramamoorthi et al. analyzed the properties of illumination cone in terms of spherical harmonics, the presence of attached shadows have been unexplained. Then they showed that when the attached shadows are clarified, the image of convex, Lambertian object can be approximated, using nine basis images with high accuracy. Firstly, the illumination cone and a few works presented then the spherical harmonic representation is explained.

The equation (2.2) is the key to producing linear lighting models that is dealing with the attached shadows. This equation describes how lighting is analogues to a convolution on the surface of the sphere. In this equation, reflectance is determined for each surface normal, v_r, by integrating the coming light from all directions weighted by kernel k. For every v_r, the kernel is rotated version of the same function and named as half-cosine function.

Since the convolutions are used, they can be analyzed in frequency domain. Fourier basis and harmonic basis are similar. Since unit vectors corresponding to the surface normals the appropriate signal-processing tools are spherical harmonics (on a sphere). These are similar to Fourier series in 2d (on a circle). These tools are examined below:

The spherical harmonics function denoted by h_{nm}. h_{nm} is the n'th order spherical harmonic and represented such as in equation (2.3). These functions represent a set of functions that form on orthonormal basis for the set of all functions on the surface of the sphere. The first part of the equation is normalization factor.

$$h_{nm}(\Theta,\Phi)=\sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}}P_{nm}(\cos\Theta)e^{im\Phi}\ , \tag{2.3}$$

where $n=0,1,\ldots,-n\le m\le n$ and P_{nm} is associated Legendre function defined as

$$P_{nm}(z)=\frac{(1-z^2)^{m/2}d^{n+m}}{2^n n!dz^{n+m}}(z^n-1)^n\ . \tag{2.4}$$

The spherical harmonics form an orthonormal basis, any piecewise, continues function on the surface of the sphere can be written as a linear combinaton of an infinite series of harmonics. Therefore, kernel k and the light l can be represented as linear combinations of the surface harmonics. Equation (2.5) is the harmonic expansion of light and equation (2.6) is the harmonic expansion of the kernel.

$$l=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}l_{nm}h_{nm} \tag{2.5}$$

$$k(u)=\sum_{n=0}^{\infty}k_n h_{n0} \tag{2.6}$$

Funck-Hecke theorem shows that convolution in the function domain is equivalent to multiplication in harmonic domain. Therefore, it can be shown that the possible reflectance of a sphere all lie near a low-dimensional linear subspace of the space of all function defined on the sphere.

$$r=k*l=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}(\sqrt{\frac{4\pi}{2n+1}}k_n l_{nm})h_{nm} \tag{2.7}$$

The Funck-Hecke theorem explained that when the reflectance function, r, the amplitude of the light l, at every order n is scaled by a factor that depends only on the convolution kernel k. This is used to determine which frequencies dominate r.

The harmonic expansions of the Lamabertion kernel k is as follows:

$$
k_n = \begin{cases} \dfrac{\sqrt{\pi}}{2} & ,\quad \text{n=0} \\ \sqrt{\dfrac{\pi}{3}} & ,\quad \text{n=1} \\ (-1)^{\frac{n}{2}+1} \dfrac{(n-2)!\sqrt{(2n+1)\pi}}{2^n(\frac{n}{2}-1)!(\frac{n}{2}+1)!} & ,\quad \text{n} \geq 2,\ \textit{even} \\ 0 & ,\quad \text{n} \leq 2\text{, odd} \end{cases} \tag{2.8}
$$

The first few coefficients are,

$$
\begin{aligned}
&k_0 = \frac{\sqrt{\pi}}{2} \approx 0.8862 \qquad \mathrm{k}_1 = \sqrt{\frac{\pi}{3}} \approx 1.0233 \qquad \mathrm{k}_2 = \frac{\sqrt{5\pi}}{8} \approx 0.4954 \\
&k_4 = -\frac{\sqrt{\pi}}{6} \approx -0.1108 \qquad \mathrm{k}_6 = \frac{\sqrt{13\pi}}{128} \approx 0.0499 \qquad \mathrm{k}_8 = \frac{\sqrt{17\pi}}{256} \approx -0.0285 \\
&(k_3 = k_5 = k_7 = 0)
\end{aligned} \tag{2.9}
$$

The energy captured by every harmonic term is calculated by the square of its respective coefficient divided by the total squared energy of the transformed function. The total squared energy in the half-cosine function is given by

$$
\int_0^{2\pi}\int_0^{\pi} k^2(\Theta)\sin\Theta d\Theta d\Phi = 2\pi \int_0^{\pi/2} \cos^2\Theta \sin\Theta d\Theta = \frac{2\pi}{3} \tag{2.10}
$$

The relative energy captured by each of the first several coefficients is shown in Table 2.1 [19]. The kernel function, the half-cosine function can be written as

$$
k(\Theta) \approx \frac{1}{4} + \frac{1}{2}\cos(\Theta) + \frac{5}{16}\cos(2\Theta) \tag{2.11}
$$

Table 2.1: Energy captured by the n'th harmonic for the Lambertian kernel. The middle row shows the cumulative energy up to n. The last row shows the energy that might be affected by the non negative light. The middle row also shows the quality of the approximation of reflectance function.

N	0	1	2	4	6	8
Energy	37.5	50	11.72	0.59	0.12	0.04
Cumulative Energy	37.5	87.5	99.22	99.81	99.93	99.97
Lower bound	37.5	75	97.96	99.48	99.80	99.90

2.3 Approximating the Reflectance Function

The lambertian kernel has most of its energy in the low-ordered terms, so it acts as a low-pass filter. As a result, any reflected function under any light can be approximated using only low-order spherical harmonic reflectances. It is, assumed that, lighting generated by a point light source at the z direction and it is a delta function. The reflectance of a sphere under illumination by a point source is calculated by a convolution of the delta function with the kernel. Appoximation of reflectance due to this point light source corresponds to 99.22 of the energy. The approximation is obtained by using the linear combination of first three harmonics. By using first order harmonics 87.5% of the energy, by using fourth order harmonic 99.81% of the energy is captured. The quality of the approximation improves with the addition of the fourth order term and deteriorates when a first order approximaiton is used. The graphical representation of energy that is captured by harmonic coefficients is in Figure 2.1 [19].

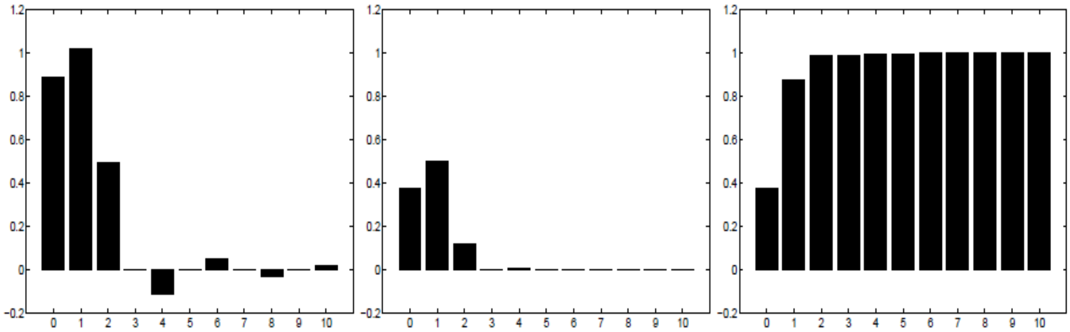


Figure 2.1: Representatoin of the first 11coefficients, the relative energy captured by each coefficients and accumulated energy, respectively.

If the direction of the point light source is different than z, the reflectance that obtained would be the same, but shifted in phase. Shifting in phase is distributed the energy over the harmonics of the same order, but overall enery, on each harmonics of the n'th order is not changed.

2.4 Generating Harmonic Reflectances

Basri et al. showed that constructing a basis to the space can be done analytically. These basis are low ordered hamonics and can be obtained by using Funck-Hecke theorem. In many cases The basis vector for the nm components of reflectances to indicate the reflectance produced by a corresponding basis vector describing the lighting h_{nm} is wanted**.** This makes relating the reflectances and lighting easy. These reflectences called harmonic reflectances and to get these reflectances the half-cosine kernel is convolved with simple harmonics. The reflectance harmonics are calculated as in equation (2.9):

$$r_{nm} = k * h_{nm} = c_n h_{nm}, \qquad \textbf{(2.12)}$$

where h_{nm} is the harmonic light and k is the kernel. To determine c_n Funck-Hecke theorem is used and finally r_{nm} is given by

$$r_{nm} = \left(\sqrt{\frac{4\pi}{2n+1}} k_n \right) h_{nm}. \qquad \textbf{(2.13)}$$

The first few given harmonics are given by

$$r_{00} = \pi h_{00} \qquad r_{2m} = \frac{\pi}{4} h_{2m} \qquad r_{6m} = \frac{\pi}{64} h_{6m}$$

$$r_{1m} = \frac{2\pi}{3} h_{1m} \qquad r_{4m} = \frac{\pi}{24} h_{4m} \qquad r_{8m} = \frac{\pi}{128} h_{8m} \tag{2.14}$$

for $-n \le m \le n$ ($r_{3m} = r_{5m} = r_{7m} = 0$).

It is useful to espress the harmonics in terms of (x, y, z) rather than (Θ, Φ).

For the angles $\Theta = \cos^{-1} z$ and $\Phi = \tan^{-1} \frac{Y}{X}$ are used to obtain the first nine harmonics in terms of (x, y, z).

$$h_{00} = \frac{1}{\sqrt{4\pi}} \qquad h_{11}^{o} = \sqrt{\frac{3}{4\pi}} y \qquad h_{21}^{o} = 3\sqrt{\frac{5}{12\pi}} yz$$

$$h_{10} = \sqrt{\frac{3}{4\pi}} z \qquad h_{20} = \frac{1}{2}\sqrt{\frac{5}{4\pi}} (2z^2 - x^2 - y^2) \qquad h_{22}^{e} = \frac{3}{2}\sqrt{\frac{5}{12\pi}} (x^2 - y^2)$$

$$h_{11}^{e} = \sqrt{\frac{3}{4\pi}} x \qquad h_{21}^{e} = 3\sqrt{\frac{5}{12\pi}} xz \qquad h_{22}^{o} = 3\sqrt{\frac{5}{12\pi}} (xy)$$

(2.15)

where superscripts e and o denote the even and odd components of the harmonics respectively.

2.5 Representing Images in Terms of Reflectances

The main aim of Section 2 is to show how to represent the set of images of objects seen under varying illumination**.**

Basri et al. constructed a simple way for this. In this way, each point of the object takes its intensity from the point on the sphere whose normal is the same. Then the intensity is scaled by the albedo of that point as in equation (2.16).

$$I(p) = \lambda r(n_x, n_y, n_z) \tag{2.16}$$

I is the image point p with surface normal $n(n_x, n_y, n_z)$, reflectance function $r(n_x, n_y, n_z)$, albedo λ and $r(n_i) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} l_{nm} r_{nm}(n_i)$

We explain the harmonic images shortly as follows. The linear space spanned by harmonic images approximates to the set of images of that object . These harmonic images named b_{nm}. Harmonic images are images of an object seen under harmonic light. These images are constructed as follows:

$$b_{nm}(p) = \lambda r_{nm}(n_x, n_y, n_z) \qquad \textbf{(2.17)}$$

b_{00} contains for every point simply the surface albedo at the point because it is an image obtained under constant light. b_{1m} contains the three components of the surface normals, scaled by albedos for every point. These images are obtained under cosine lighting centered at the three main axis. The higher order harmonic images contain polynomials of the surface normals scaled by the albedo.

By combining equations (2.14) and (2.17) they obtain the harmonic images as follows.

$$b_{00} = \frac{1}{\sqrt{4\pi}}\lambda \qquad b_{11}^{o} = \sqrt{\frac{3}{4\pi}}\lambda.*n_y \qquad b_{21}^{o} = 3\sqrt{\frac{5}{12\pi}}\lambda.*n_{yz}$$

$$b_{10} = \sqrt{\frac{3}{4\pi}}\lambda.*n_z \qquad b_{11}^{e} = \sqrt{\frac{3}{4\pi}}\lambda.*n_x \qquad b_{21}^{e} = 3\sqrt{\frac{5}{12\pi}}\lambda.*n_x n_z$$

$$b_{20} = \frac{1}{2}\sqrt{\frac{5}{4\pi}}\lambda.*(2n_{z^2} - n_{x^2} - n_{y^2}) \qquad b_{22}^{e} = \frac{3}{2}\sqrt{\frac{5}{12\pi}}\lambda.*(n_{x^2} - n_{y^2})$$

$$b_{22}^{o} = 3\sqrt{\frac{5}{12\pi}}\lambda.*n_x n_y \qquad \textbf{(2.18)}$$

where p_i denote the i'th object point, λ denote a vector of the object's albedos, λ_i is the albedo of p_i. Similarly, n_x, n_y and n_z denote the three vectors of the same length that contain the x, y and z component of the surface normal of p_i. In addition, n_{x^2} denote a vector $n_{x^2,i} = n_{x,i} n_{x,i}$ and the definitions of n_{y^2}, n_{z^2}, n_{xz}, n_{yz}

and n_{xy} is similar to this one. Then the intensity image can be obtained accordıng to equation (2.19).

$$I_i = \sum_{i=-n}^{\infty} \sum_{m=-n}^{n} l_{nm} b_{nm}(p_i) \tag{2.19}$$

From the equation (2.19), we can say the dependecies of the accuracy of the approximation of spherical harmonics to images. The shape and the albedo of the object is very important. If the surface normals of the object differs from the normals of the sphere, the accuracy of approximation is decreased. It is also decreased with the viewing direction, because this affects the distribution of normals. Also, because of scaling of each pixel property of albedo, it affects the accuracy of this approximation.

2.6 Recognition

The recognition process is done under some conditions such as known pose of face, and unknown identiy and lighting conditios. For recognition a new image is compared to each image in turn that are in the database. The distance between the image and the nearest image in the database is used for comparison. The linear subspace or a subset of the linear supspace can be used while calculating distance.

The most common and appropriate way for face recognition is to compare a new image to the linear subspace of images that corresponds to a image that is in the database. To do this, harmonic basis images of each model is produced as in equation (2.18**)**. Given a new image I, a vector α that minimizes $\| B\alpha - I \|$ is found. B is the set of basis images with size $r * d$, d is the number of pixels in the image and r is the number of basis images used. Every row of B contains one spherical harmonic image and the rows of B form a basis for the linear subspace. These images form a basis for the linear subspace. The QR decomposition is applied to B, to obtain an orthonormal basis. The size of Q is same as B with orthonormal columns, and R is a $r * r$ matrix so $QR = B$ and QQ^T is an $r * r$ identity matrix. Then the distance between I and space spanned by B is calculated as $\| QQ^T I - I \|$.

3. SPHERICAL HARMONICS AND STATISTICAL MODEL BASED FACE RECOGNITION USING ONE TRAINING IMAGE PER PERSON

3.1 Introduction

As in Section 2, the set of images of a convex Lambertian object obtained under unknown lighting conditions can be approximated by a nine-dimensional linear subspace. The basis images can be recovered from just one image taken under arbitrary light and fixed pose. For using this method, basis images that can be rendered from a 3D scan of a face or can be estimated by applying PCA to a number of images of the same object under varying illumination required because they span all the illumination space for each face.

Lee et al. in [22], suggested an effective approximation of this basis images by nine single light source images of a face and Wang et al. in [46] suggested an illumination modelling and normalization method for face recognition. All of these methods have limited applicability, because they need a number of trainig images or 3D scans of the subjects in the database. Samaras & Zhang, introduce a method that requires only one training image for each subject with no illumination limitations by using spherical harmonics and statistical methods. Because of using statistical model, spherical harmonics and recognition scheme similar to PCA, this method is a hybrid face recognition method as explained in Section 1.6.3. This method is based on [19, 20] and the method consist of three steps.

1. Statistical model computation
2. Training
3. Testing

In statistical model computation, for each of the spherical harmonics a statistical model is computed by using a bootstrap set. The aim of this statistical model is to learn the probability density functions for the basis images. The most heaviest part of the method is the construction of the bootstrap set and the statistical model computation. But the bootstrap set and the resulting probability functions obtained only once. When new images are added to database there is no necessary modifications to do in this part, but if the modifications are done, the scalability of the system may increase. In training step, the results of the statistical model computation part are used. When a new image is given, firstly the weight of the basis images estimated, than maximum a posteriori method used for recovering the basis images. For each face in the training set, basis images are recovered. The basis images spans the illumination space for this face. The faces in the training set do not have to be in bootstrap set. Only necessity is similar statistical characteristics for the images in bootstrap set and the training images. In testing part, the recognition scheme that is explained in Section 2.6 is used.

3.2 Getting Harmonic Images

In this step, the harmonic images are calculated. For remembering how to obtain reflectance, intensity and spherical harmonics, the terms are shortly explained below:

When L shows the distant lighting distribution and the cast-shadows and near-field illumination is neglecting, the irradiance E, function of the surface normal n , can be represented as below.

$$E(n) = \int L(w)(n.w)dw \tag{3.1}$$

To obtain image intensity (radiosity), E (irradiance) must be scaled by the surface albedo λ to find the radiosity I as in below.

$$I(\rho, n) = \lambda(\rho)E(n) \tag{3.2}$$

The spherial harmonics are a set of functions and analogue on the sphere to the Fourier basis on the line or circles as explained before. Basri & Jacobs, and also Ramamorthi & Hanrahan has independently showed that the combination of the first nine-spherical harmonics are approximates to E for Lambertian surfaces. The equations that explain how to obtain the harmonics is same in equation (2.15).

According to (2.15) the image intensity of a point p with surface normal n(n_x, n_y, n_z) and albedo λ can be computed by replacing x,y,z with n_x, n_y, n_z. Then the way of obtaining spherical harmonics is applying the equation set in (2.18).

3.3 Recovery of Basis Images from 2D Statistics for Fixed Pose

3.3.1 Statistical models of basis images

$$i(x) = b(x)^T \alpha + e(x, \alpha) \tag{3.3}$$

where at pixel position x the pixel intensity i(x) is the weighted combination of the basis images b(x) plus an error term e(x).

Equation (3.3) is the main equation of this method. My thesis is based on this equation.

$$I = B^T \alpha + E \tag{3.4}$$

α is the set of illumination coefficients and 9*1 dimensional vector. I is a d dimensional image vector and E is a d dimensional vector showing the error term. B is a $9*d$ dimensional matrix of basis images, the columns of B are the vectors $\{b(x)\}_{x=1}^{d}$, in other words the basis images that are d dimensional.

The aim of the statistical models is to learn the probability density functions (pdf) for B and E. The pdfs of both B and E are assumed to be Gaussian distributions of unknown means and covariances. In this method they can be estimated from the basis images. From these basis images, the sample mean vectors $\mu_b(x)$ and the sample covariance matrixes $C_b(x)$ are computed. For the ith (i=1..9) basis, the component of the covarince matrix $C^i{}_{m,n} = \mathrm{cov}(b^i_m, b^i_n)$ where b^i_j representing the jth (j=1….d) pixel vector consisting of the ith pixels of the N bootstrap faces. $C_b(x)$ is 9*9 dimensional matrix, while $\mu_b(x)$ is a 1*9 vector.

To estimate the statistics of the error term, 2d images of each face are illuminated with different illuminants and estimated their coefficients. After illuminating each face, least-squares method is applied to estimate the coefficients α. Then the sample mean $\mu_e(x,\alpha)$ and sample variance $\delta_e^2(x,\alpha)$ of the error term are calculated by using equation (3.5)

$$E = I - B^T \alpha \qquad \textbf{(3.5)}$$

The error term shows the difference between the Lambertian reflectance and the errors of the low-dimensional approximation which cause the estimates of the coefficients to diverge from the true values. Because of mean and variance are functions of α, learning the statistics of the error term from a finite number of samples, is impossible. Given a new image with illumination coefficients α_{tra}, the mean and the variance of the error term at α_{tra} need to be calculated using kernel regression.

In this method, the statistical method is learned for evey pixel and the spherical harmonics at different pixels assumed to be independent of one another. In the recovering part, this assumption is considered.

3.3.2 Recovering the basis images

When a new image is given, under any illumination, first the coefficients α are estiamated than the error term is estimated using the coefficients. Lastly, the basis images recovered at each x by computing maximum a posteriori (MAP) estiamate,

$$b_{MAP}(x) = \arg\max_{b(x)}(P(b(x) \mid i(x))) \qquad \textbf{(3.6)}$$

3.3.2.1 Estimating harmonic coefficients

For estimating the unknown spherical harmonic coefficiens α for the new face image, kernel regeression is used. Firstly, all the K bootstrap images $\{J_k\}_{k=1}^K$ along with their spherical harmonic coefficients $\{\alpha_k\}_{k=1}^K$ are loaded. Given a new face image, the coefficients can be estimated using equation set (3.7) .

$$\alpha_{tra} = \frac{\sum_{k=1}^{K} w_k \alpha_k}{\sum_{k=1}^{K} w_k} \qquad \text{where } w_k = \exp[-\frac{1}{2}(D(i,J_k)/(\sigma_k)^2]$$

$$D(i,J_k) = \| i - J_k \|_2 \qquad \textbf{(3.7)}$$

where σ_k, is the width of the kth Gaussian kernel which controls the influence of J_k. σ_k is precomputed in a way such that 10 percent of bootstrapimages are $1*\sigma_k$ at each σ_k.

3.3.2.2 Estimating the error term

Both the sample mean and sample variance are functions of α. The error term E(α) estimated by using kernel regression. The mean and the variance of the error estimation $e(x,\alpha_{tra})$ at α_{tra} will be interpolated from the known mean and variance of the error term of coefficients $\{\alpha_k\}_{k=1}^{K}$ which have been calculated. We use equation (3.8) for mean of error-term and equation (3.9) for variance of error.

$$\mu_e(x,\alpha)=\frac{\sum_{k=1}^{K} w_k \mu_e(x,\alpha_k)}{\sum_{k=1}^{K} w_k} \qquad \text{where} \quad w_k=\exp[-\frac{1}{2}(D(\alpha,\alpha_k)/(\sigma_k)^2]$$

$$D(\alpha,\alpha_k)=\|\alpha-\alpha_k\|_2 \qquad \textbf{(3.8)}$$

$$\sigma_e^{\,2}(x,\alpha)=\frac{\sum_{k=1}^{K} w_k \sigma_e^{\,2}(x,\alpha_k)}{\sum_{k=1}^{K} w_k} \quad \text{where } w_k=\exp[-\frac{1}{2}(D(\alpha,\alpha_k)/(\sigma_k)^2]$$

$$D(\alpha,\alpha_k)=\|\alpha-\alpha_k\|_2 \qquad \textbf{(3.9)}$$

where, σ_k, is the width of the kth Gaussian kernel which controls the influence of α_k.

3.3.2.3 Computing the basis images

Equations in (2.18) are used for calculating the basis images. The results of estimation of coefficients, the mean and the variance of the error term and also mean and covariance of basis images are used for recovering the basis images of a given image. As said before, the basis images are recovered by computing MAP estimate but calculation of this estimation is hard directly. So, using Bayes' rule,

$b_{MAP}(x)=\arg\max_{b(x)}(P(b(x)\,|\,i(x)))$, it turns into:

$(P(b(x)\,|\,i(x))P(i(x))=(P(i(x)\,|\,b(x))P(b(x))$

The given image i(x) and coefficients α are known , $P(i(x))$ is constant. Then,

$$b_{MAP}(x) = \arg\max P_b(x)((i(x) \mid b(x))P(b(x))) \quad \textbf{(3.10)}$$

In this term $i(x)$ $P_b(b(x))$ is the Gaussian probability density function that learned previously, $P(i(x) \mid b(x))$ can be computed from equation (3.4). Thus, $i(x)$ is a random variable with Gaussian pdf of mean $b(x)^T \alpha + \mu_e(x,\alpha)$ and variance σ_e^2. Then,

$$b_{map} = \arg\max_{b(x)} \left[Gauss(b(x)^T \alpha + \mu_e, \sigma_e^2) \times Gauss(\mu_b(x), C_b(x) \right] \quad \textbf{(3.11)}$$

Using log probability and ignoring constant terms equation (3.11) turns into equation (3.12).

$$\arg\max_{b(x)} L = \frac{-1}{2}\left(\frac{i - b^T \alpha - \mu_e}{\sigma_e}\right)^2 - \frac{1}{2}(b - \mu_b)^T C_b^{-1}(b - \mu_b) \quad \textbf{(3.12)}$$

After taking the derivative of this equation with respect to b(x) to find the maximum probablity, the equation becomes as follows:

$$\frac{\partial L}{\partial b(x)} = \frac{-2}{{\sigma_e}^2}(i - b^T \alpha - \mu_e)\alpha + 2Cb^{-1}(b - \mu_b) = 0 \quad \textbf{(3.13)}$$

Rearranging the equation (3.13) the equations (3.14) , (3.15) and (3.16) are obtained.

$$A * b_{map} = T \;, \quad \textbf{(3.14)}$$

where $A = \frac{1}{{\sigma_e}^2} \alpha\alpha^T + Cb^{-1}$ and **(3.15)**

$$T = \frac{(i - \mu_e)}{{\sigma_e}^2} \alpha + Cb^{-1} \mu_b \quad \textbf{(3.16)}$$

Recovery of basis image process is done for each pixel seperately. For this reason, A is a $d*d$ matrix, b_{map} is $d*d$ a matrix and T is also $d*d$ matrix, i, μ_e and σ_e^2 corresponds to the related pixel of image and error term, all of them is 1 dimensional. α is a $d*1$ dimesional vector and C_b^{-1} and μb are $d*d$ and $d*1$ dimensional vectors respectively. d is the number of basis images that are used.

3.4 Recognition

In recognition part, the recognition model that is explained in Section 2.6 is used. The closest training face image to testing one is found using the combination of basis images of training images. B is the set of basis images with size $r*d$, d is the number of pixels in the image and r is the number of basis images used. QR decomposition is applied to B, to obtain an orthonormal basis. The distance is computed from the test image and space spanned by B as

$$\| Q\, Q^T\, I - I \| \tag{3.17}$$

4. FACE RECOGNITION USING HARMONIC IMAGES

Work of Samaras et al. in [19, 20] that is explained in Section 3 is implemented with some differences and the experimental results are presented in the next sections. The differences between the implemented one and the explained one in Section 3 are in recovering the basis images step. In estimation of harmonic coefficients, mean and variance of the error term, a different method is used. The testing and the training steps are almost same.

Because of used database as bootstrap set, the images need preproccessing. The used steps are explained during the Section 4.1.

4.1 Preprocessing

As we describe in Section 1, face alignment is an important part of face recognition systems. Face alignment considerably affects the recognition results. So, we applied face alignment to all images we use in the experiments. Each face image is aligned with using their eye and lip positions and then cropped. An example of aligned and cropped image can be seen in Figure 4.1.

Initially we do not have albedo and surface normal information of images that we use in bootstrap set, but for each face (50 face images) we have 2D and 2.5D images. So, we can calculate albedos and the surface normals of faces to be able to obtain the spherical harmonics and basis images by using 2D and 2.5D images of faces. However, 2.5D images have invalid values at some x,y coordinates. To construct a true surface of face images, surface reconstruction is applied to 2.5D images firstly.

Figure 4.1: The intensity image and the aligned and cropped face image.

4.1.1 Surface reconstruction

Because of 2.5D images have some invalid values, the result of process that we find surface normals are not correct. For surface reconstruction regularization, a theoritical approach to solve ill-posed problems, is used [48]. In regularization a unique solution is found by imposing constraints on the solution.The reconstructed surface is obtained by finding the function f that minimizes the eqution (4.1).

$$E_m(\lambda, f) = \iint (f-d)^2 dxdy + \lambda \iint (f_x^2 + f_y^2) dxdy \qquad \textbf{(4.1)}$$

E_m is an energy functional associated with a membrane. In the function first term ($\iint (f-d)^2 dxdy$) is the measure of closeness of f to the data(range)d, and the second term($\iint (f_x^2 + f_y^2) dxdy$) is the measure of smoothness. The detailed information can be found in [48].

4.1.2 Getting surface normals

The surface Z, can be thought of as a function of the (x,y) coordinates on the image.

Z=Z(x,y). **(4.2)**

The surface slopes can be computed by taking the x and y partial derivatives of the vector , $[x, y, Z(x,y)^T]$.

$[0,1,\frac{\partial z}{\partial x}]^T$ x derivative

$[1,0,\frac{\partial z}{\partial y}]^T$ y derivative

For taking the derivative of the image we use filters. This filters can be shown in Figure 4.2. The other filters such as sobel filters are tried but with this one the best result is obtained.

-1	0	1

(a)

-1
0
1

(b)

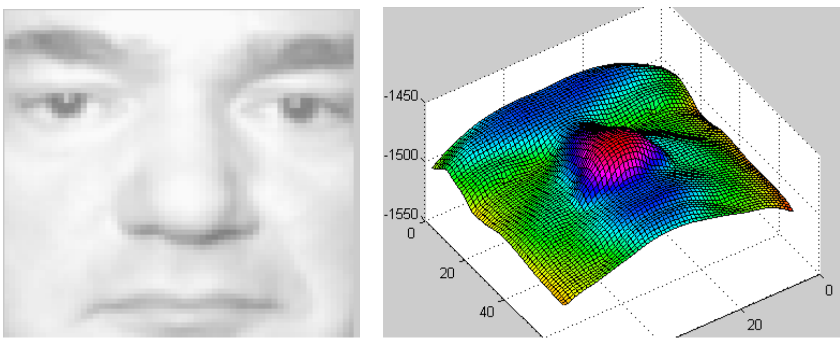
Figure 4.2: Image filters for obtaining surface normals. (a) filter used for taking x derivative. (b) filter used for taking y derivative.

x derivative of the vector corresponds p, and y derivative of the vector corresponds to q.

The surface normals of an image can be obtained by using equation (4.3).

$$n = 1/(\sqrt{1+p^2+q^2})[-p,-q,1]^T \qquad \textbf{(4.3)}$$

In Figure 4.3 , an intensity image and the surface of this image can be seen. The Z component of 2.5D image is used while the surface is drawn and this is also used in getting normals.



(a) (b)

Figure 4.3: The intensity image of one subject and the reconstructed surface of this subject.

4.1.3 Getting albedos

Previously, we explianed that the intensity image I, can be represented as in equation (3.2). From equation (3.2) the albedo of an image can be found by

$$\lambda(\rho) = \frac{I(\rho, n)}{E(n)} \tag{4.4}$$

For this equation we firstly find the reflectance of that image as in the equation set (4.5)

$$E(x, y) = \frac{\rho}{\sqrt{1 + p^2 + q^2}} i^T [-p, -q, 1]$$

$$\max\left\{0, \frac{\rho}{\sqrt{1 + p^2 + q^2}} i^T [-p, -q, 1])\right\} \tag{4.5}$$

For each image we firstly calculate E and then we find albedo with using E and I. The database that we use consists the illumination direction for all images, so we can easily obtain the albedos of each face. In Figure 4.4 three faces and albedos of each faces showed.

Figure 4.4: The intensity images and albedos of three subjects

4.2 Statistical Models of Harmonic Images

After the images are aligned and cropped, the component of the surface normals (n_x, n_y, n_z) of the faces and also the albedo of the faces are obtained. Then, using the equation set (2.18), the harmonic images for each face image in the bootstrap set are calculated. For statistical model calculation, firstly the sample mean vector and the sample covarince matrixes are calculated as explained in Section 3.3.1. Figure 4.5 displays the calculated basis images of one subject and Figure 4.6 displays the mean of the statistical models of the basis images computed from the bootstrap set that we have used in experiments.

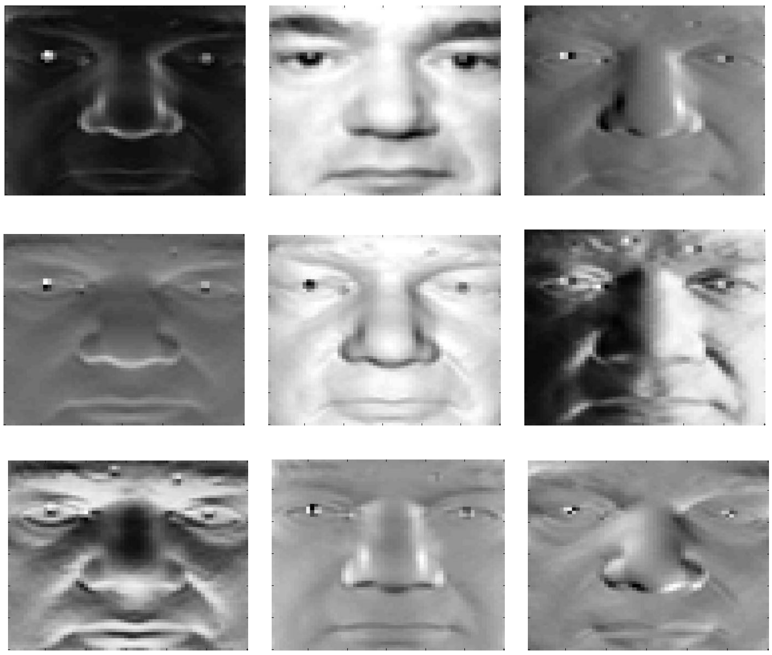


Figure 4.5: The rendered basis images of one subject, computed using 2d and 3d images of subject. b00, b10, b11e, b11o, b20, b21o, b21e, b22e, b22o respectively.

Figure 4.6: The mean images of the statistical models computed from the rendered basis images. b00, b10, b11e, b11o, b20, b21o, b21e, b22e, b22o respectively.

4.3 Recovery of Basis Images from 2D Statistics for Fixed Pose

When we used kernel regression as in equations (3.7), (3.8) and (3.9) for estimating the harmonic coefficients and the statistical properties, mean and the variance of the error term of a given image under any illumination, the results was not as expected. So, we tried another algorithm for estimating the harmonic coefficients and the statistical properties. For each of the image in bootstrap set, we can calculate the harmonic coefficients using equation (4.6) by ignoring the error term.

$$\alpha = \text{inv}(\text{B*B}^T)\text{*(B*I)} \tag{4.6}$$

For estimating the harmonic coefficients and the statistical properties of the error term, firstly we load all K bootstrap images $\{J_k\}_{k=1}^{K}$. After calculating the basis images of them, we illuminated each bootstrap image with using 64 different illumination conditions (the illumination directions are same as in YaleB database) $\{I_n\}_{n=1}^{K*64}$. For each image in $\{I_n\}_{n=1}^{K*64}$, we calculated the error term and the harmonic coefficients of them, $\{E_n\}_{n=1}^{K*64}$ and $\{\alpha_n\}_{n=1}^{K*64}$ using equations (3.5) and (4.6) respectively. When a new image is given, $\{E_n\}_{n=1}^{K*64}$ and $\{\alpha_n\}_{n=1}^{K*64}$ are used for estimation of harmonic coefficients and the statistical properties of error term. The distances between the given new image and $\{I_n\}_{n=1}^{K*64}$ are calculated firstly and then the most nearest m ones to training image of $\{I_n\}_{n=1}^{K*64}$ are chosen and stored in T. Then using equation (4.7), the harmonic coefficients of the new image are estimated with a similar method that is explained in Section 3.3.2.1.

$$\alpha_{new} = \frac{\sum_{i=1}^{m} w_{T_i} \alpha_{T_i}}{\sum_{i=1}^{m} w_{T_i}} \qquad \textbf{(4.7)}$$

where $w_{T_i} = \exp[-\frac{1}{2}(D(I_{new}, I_{T_i})]$ and D is the Euclidean distance. T includes the indexes of the most nearest images to given new image and m represents the number of used nearaest images .

Use ,

M = mod (T-1 , ni) + 1 **(4.8)**

to find the indexes of the illuminations that the illumination of new image most resembles to. Here *ni* is the illumination number.

Then, for finding the mean and the variance of the error term use $\{E_n\}_{n=1}^{K*64}$**,** T and equation (4.8). Mean and the variance of the error term of a pixel are scalar values and calculated using the equations in (4.9) .

$$\mu_e(x) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{K} E_{T_i}^{j}}{m*K} \qquad \sigma_e{}^2(x) = \frac{\sum_{i=1}^{m} \sum_{k=1}^{K} (\sigma_e^2)_{M_i}^{j}(x)}{m} \qquad \textbf{(4.9)}$$

Where M includes the indexes of the most nearest illuminations to illumination of the given new image and m represents the number of used nearaest images.

Mean of error term corresponds to mean of the errors of M'th (*M* is a vector) illuminations that the illumination of test image is most resemble to. Variance of the error term corresponds to mean of the variances of the M'th (*M* is a vector) illuminations that the illumination of test image is most resemble to. In Figure 4.7 when a new image is given, the algorithm of estimating the harmonic coefficients and the statistical properties of the error term is presented.

1. Load all K bootstap images $\{J_k\}_{k=1}^{K}$, J is a d dimensional vector
2. Calculate their basis images (Equation 2.18)
3. Illuminate eack K bootstrap image with using 64 different illumination and save them as $\{I_n\}_{n=1}^{K*64}$, I_n is a d dimensional vector
4. For each image in $\{I_n\}_{n=1}^{K*64}$ calculate the error term using corresponding basis images(Equation 3.5) and save them as $\{E_n\}_{n=1}^{K*64}$, E is a d dimensional vector
5. For each image in $\{I_n\}_{n=1}^{K*64}$ calculate the harmonic coefficients (Equation 4.6) using corresponding basis images and save them as $\{\alpha_n\}_{n=1}^{K*64}$, α is a 9 dimensional vector.
6. Given a new image I_{new}, calculate the distances between I_{new} and $\{I_n\}_{n=1}^{K*64}$
7. Choose the most nearest m ones of I_n and load the indexes in T
8. Estimate the harmonic coefficients of I_{new} (Equation 4.7)
9. Estimate the mean error and variance of error as in (Equation 4.8, Equation 4.9)

Figure 4.7: Algorithm of estimating the statistical properties of error term and harmonic coefficients.

After estimating the statistical properties both the error term and the harmonic images, using equation (3.14), (3.15) and (3.16), the harmonic basis images of the given face image is obtained. While basis image recovery and the calculation of statistical properties of error term, the equations must be used for per pixel. The pseudo-code of the recovering basis images of a given new image can be seen in Figure 4.8.

```
test, given new image;
coord, harmonic coefficients of Test;
for each pixel in Test,m
    a, the vector that includes the mth pixels of all basis images
    t, reshape 'a' as the rows indicate basis and the columns
       indicate subjects for mth pixel.

    mb = mean(t) ,9 dimensional
    Cb = cov(t'), 9*9 dimensional

    error, the vector that holds the error term of all images in
           {I_n}_{n=1}^{K*64} at mth pixel
    T, reshape 'error' as the rows indicate the type of
       illuminaiton, columns indicate subjects for mth pixel

    me = mean(t(k)), mean of the errors of kth(k is a vector)
         illuminations that the illumination of test is more
         resemble to
    vare = mean(var(t(k)), mean of the variances of the kth(k is a
           vector) illuminations that the illumination of test
           is more  resemble to

    A = 1/vare * coord * coord' + inv(Cb)
    T = (test(m)-me)/vare * coord + inv(Cb)*mb
    basis(m) = inv(A)*T, includes the basis of mth pixel

end
```

Figure 4.8 Pseudo-code of recovering the basis images for a given new image

4.4 Recognition

In recognition part, the recognition model that is explained in Section 2.6 is used . We choose randomly one image per person from subset1 and subset2 to use in training part and the remaining parts of susbet1 and subset2 are used for testing as well as subset3 and subset4.

5. EXPERIMENTS AND RESULTS

We used a database that is obtained from University of Notre Dame in United States. The database is part of University of Notre Dame Biometrics Database Distribution and named as Collection D. It contains the 2D and 2.5D (range) images of the same person that are captured in the same time and same environment. It has 277 subjects and 1096 images. The size of each image is 640*480. The 50 subject have chosen randomly and used as bootstrap set for this experiment and 10 subject have chosen randomly to be used in testing and training sets.

For testing the method, that we explained in Section 4, the images in the database have to be preprocesing firstly. The images are aligned and then cropped as said before in Section 4, the new size of each face image is 68*59.

All images in bootstrap set are illuminated using 64 different illumination directions and the obtained 50*64 images are used for estimating the harmonic coefficients and the statistic probabilities of the given new image in testing. We illuminated each face as in YaleB, because the database provides images that sample perfectly the whole illumination space and it is a testing standard for variable illumination recognition methods. So, the dataset that we used in the experiments, samples the illumination space sufficiently. We turn the elevetion and azimuth values into (x, y, z) coordinate system. Then we illuminated all faces that we used in bootstrap set with these 64 different illuminations. So, we have 64 images per subject. In Figure 5.1, you can see the images of one subject that is illuminated as in YaleB database.

We used in testing 45*10 face images for 10 subjects in a single pose. Each iamage of the subjects illuminated under different directional lights. The seperation of these images that are used in testing is done according to Table 5.1. The subset4 includes the extreme lighting.

Figure 5.1: One bootstrap image that is illuminated from 64 different illumination directions

Table 5.1: The seperation of test images according to YaleB database

Subsets	Subset1	Subset2	Subset3	Subset4
Number of Images	70	120	120	140
Illumination direction	0-12	13-25	26-50	51-77

5.1 Basis Images Recovery

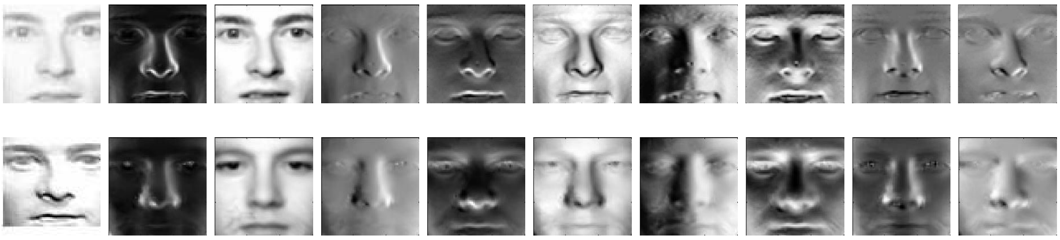


Figure 5.2: The recovered basis images(bottom) compared to the rendered ones (top) for an individual subject in the bootstrap set.

Figure 5.2 shows the comparison of the rendered basis images and the recovered basis images of one subject. The images obtaining the harmonic basis images are diferent each other. The recovered basis images are calculated using the image in database of Notre Dame University and the rendered basis images are calculated using the image of database that includes the synthetic images. The first images (top and bottom) in Figure 5.2, are different as we have seen. In our recovery process, we derive the nine basis images and use the most nearest image from the bootstrap set while calculating the statistical properties of error term.

The recovered basis images are very important, because the recognition algorithm depends on the recovered basis images. The recovered basis images of a person under different illumination directions are shown in Figure 5.3. The resulting basis images are very close as seen in Figure 5.3 except under extreme lighting. The accuracy of this method depends on the similarity of basis images under different illuminaiton directions of one subject.

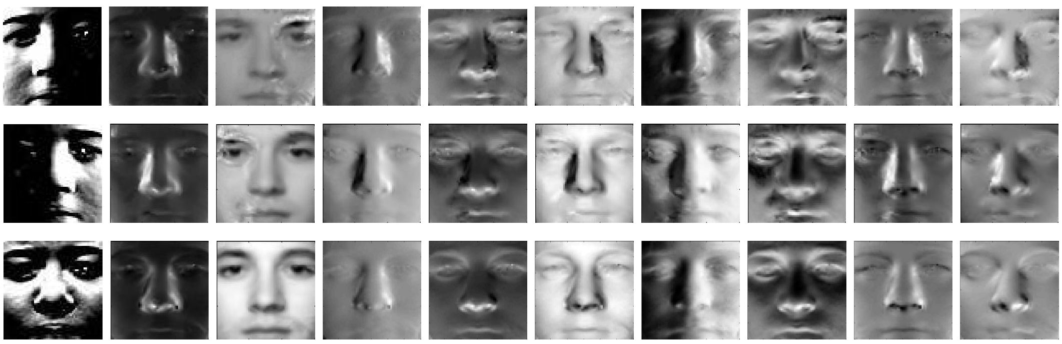


Figure 5.3: The basis images recovered from images of the same subject under various single directional illumination. The first column is the image we used for the recovery, the following ones are the basis images.

The harmonic images capture more than 99 percent of the Lambertian reflectance energy [20], the weighted combination of the 9 basis images will be very close to the real images. In Figure 5.4, the weighted combination of basis images in Figure 5.2 is presented respectively, calculated according to equation (3.4) with regarding the error term.

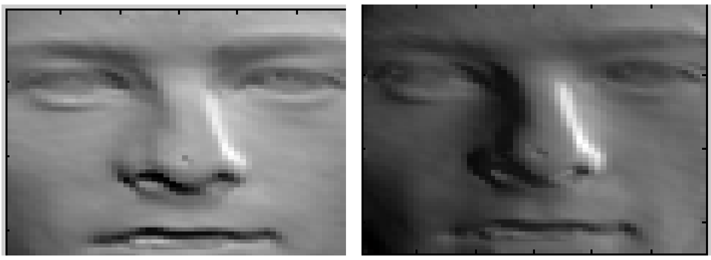


Figure 5.4: The recovered basis images using equation (3.5).

5.2 Recognition

To check the accuracy of estimation of the harmonic coefficients, we synthesized 200 images with various set of coefficients. From the synhesized images, we estimated the harmonic coefficients using equation (3.7). The result of this simple experiment showed that the two sets of harmonic images coefficients and the estimated ones are very close. The mean of the measured estimation error is 0.0767.

We experimented this algorithm on the synthesized images. As we said before in Section 4, we chose the training images from subset1 and subset2 randomly, and the remaining images are used as testing images. We repeated the process 10 times and calculate the average recognition rates as in Table 5.2.

Table 5.2: Recognition results

	Subset1	Subset2	Subset3	Subset4
Recognition rate	97.4	95.2	90.4	87.3

6. CONCLUSION AND FEATURE WORK

In this thesis, a review on spherical harmonic based face recognition algorithms is given and an algorithm that includes both spherical harmonics and statistical models is investigated [22]. A small change have done in estimation of harmonic coefficients and the statistical properties of the error term when a new image is given. We have seen that basis images can be recovered by using statistical models from a single image taken under varying illumination conditions. During the training part, only one image is used per subject, so new subjects can be added to training set easily and as we explained in Section 1, this situation is an advantage in face recognition systems. The results we had are similar to results in [22].

The different databases can be found both for bootstrap and the training and testing part. If we can find the appropriate databases for bootstrap set and the training and testing sets (the statistical properties of the databases must be very similar) for bootstrap and testing, we will plan to experiment the algorithm under more larger databases. We will also plan to develop an algorithm that solve pose variation problem using spherical harmonic images and to try different tarining methods. The experiments are done using images under one directional illumination, it will be experimented by using images that are illuminated from multiple directions.

REFERENCES

[1] Jain, A. K. and Li, **S. Z.,** 2005. *Handbook of Face Recognition*, Springer-Verlag New York, Inc., Secaucus, NJ.

[2] Zhao,W., Chellappa, R., Phillips, P. J., Rosenfeld, A., 2003. Face Recognition: A Literature Survey, *ACM Computing Surveys*, v.35 n.4, pp.399-458

[3] Okada, K., von der Malsburg, C., 2002. Pose-Invariant Face Recognition withParametric Linear Subspaces, *In Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition(FGR'02)*, 20-21 pp.71-76

[4] Gross, R., Matthews, I., Baker, S., 2002. Eigen Light-Fields and Face Recognition Across Pose, *In Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition(FGR'02)*, pp.3-9.

[5] Martinez, A., 2002. Recognizing Imprecisely Localized, Partially Occluded and Expression Variant Faces from a Single Sample per Class. *IEEE Trans. PAMI*, vol. 24, no. 6, pp.748-763.

[6] Kurita, T., Pic, M., Takahashi, T., 2003. Recognition and Detection of Occluded Faces by a Neural Network Classifier with Recursive Data Reconstruction, *In IEEE Conference on Advanced Video and Signal Based Surveillance*, 21-22 July 2003, Miami, FL, USA, pp.53.

[7] Sahbi, H., Boujemaa, N., 2002. Robust Face Recognition Using Dynamic Space Warping, *Biometric Authentication, LNCS 2359*, Springer-Verlag, Berlin, Heidelberg, pp.121-132.

[8] Lanitis, A., Taylor, C. J., Cootes, T. F., 1995. Automatic Face Identification System Using Flexible Appearance Models, *Image Vis. Comput. 13*, pp.393-401.

[9] Donato, G., Bartlett, M. S., Hager, J. C., Ekman, P., Sejnowski, T. J., 1999.Classifying Facial Actions, *IEEE Trans. PAMI*, vol. 21, no. 10, pp.974-989.

[10] Tian, Y., Kanade, T., Cohn, J. F., 2001. Recognition Action Units for Facial Expression Analysis, *IEEE Trans. PAMI*, vol. 23, no. 2, pp.97-115.

[11] Liu, Y., Schmidt K. L., Cohn, J. F., Weaver, R. L., 2002. Facial AsymmetryQuantification for Expression Invariant Human Identification, *In Proceedings of theFifth IEEE International Conference on Automatic Face and Gesture Recognition(FGR'02)*, 20-21 May 2002, Washington D.C., USA, pp.208.

[12] **Abate A.F., Nappi , M., Ricco, D, Sabatino, G.,** 2006. A survey: 2D and 3D face recognition, ScienceDirect Pattern Recognition Letters 28 pp.1885-1906

[13] **Gao, Y., Leung, M. K. H.,** 2002. Face Recognition Using Line Edge Map, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, v.24, n.6, pp.764-779.

[14] **Belhumeur, P. N., Hespanha, J., Kriegman, D.,** 1997. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, *IEEE Trans. PAMI, Special Issue on Face Recognition*, 19(7), pp.711—20.

[15] **Sim, T. and Kanade, T.** 2001. Combining Models and Exemplars for Face Recognition: An Illuminating Example, Proc. CVPR Workshop Models versus Exemplars in Computer Vision

[16] **Georghiades, A.S., Kriegman, D.J., Belhumeur, P.N.,** 1998. Illumination cones for recognition under variable lighting: Faces, *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.52.

[17] **Romdhani, S. , Blanz, V.and Vetter**, **T.,** 2002. Face Identification by Fitting a 3D Morphable Model using Linear Shape and Texture Error Functions *Proc. ECCV* (IV:3).

[18] **Li**, **S. Z., Chu, R., Liao, S., Zhang, L.,** 2007. Illumination Invariant Face Recognition Using Near-Infrared Images, *IEEE Trans. PAMI*, vol. 29, no. 4, pp.627-639.

[19] **Basri, R., Jacobs, D. W.,** 2003. Lambertian Reflectance and Linear Subspaces, *IEEE Trans. PAMI*, vol 25, no. 2, pp.218-233.

[20] **Ramamorthi, R. ,** 2002 . Analytic PCA Construction for Theoretical Analysis of Lighting Variability in Images of a Lambertian Object, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 10.

[21] **M. D'Zmura,**1991, Shading Ambiguity, *Reflectance and Illumination, Computational Models of Visual Processing.*

[22] **Zhang, L. and Samaras, D.,** 2003. Face recognition under variable lighting using harmonic image exemplars, *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conferencen,* vol 1, pp.19-25

[23] **Zhang, L. and Samaras, D.,** 2006. Face Recognition from a Single Training Image under Arbitrary Unknown Lighting Using Spherical Harmonics, IEEE Transaction on Pattern Analysis and Machine Intelligence(PAMI), Vol.28, No.3, pp. 1-8

[24] **Zhang ,L. and Samaras, D.**, 2004 Face Synthesis and Recognition from a Single Image under Arbitrary Unknown Lighting Using a Spherical Harmonic Basis Morphable Model, *Biometric Authentication, vol. 1,* pp. 10-23

[25] **Rara H., Elhabian S., Starr T. and Farag A**.,2008. A Statistical Model Combining Shape and Spherical Harmonics For Face Reconstruction and Recognition, *Proceedings of IEEE, CIBEC'08*

[26] Qing L. , ShanS., Gao W., Du B.2008. Face Recognition Under Generic Illumination Based on Harmonic Relighting, *Proceedings of the 2008 IEEE, CIBEC'08*

[27] Yue Z., Zhao W., Chellappa R., 2007. Pose-Encoded Spherical Harmonics for Robust Face Recognition Using A Single Image, *EURASIP Journal on Advances in Signal Processing* Vol.2008 Article ID 748483

[28] Zhang ,L. and Samaras, D., 2004. Pose Invariant Face Recognition under Arbitrary Unknown Lighting Using Spherical Harmonics, *Biometric Authentication, vol. 1*, pp. 10-23

[29] Tan, X.,Chen, S., Zhou, Z. and Zhang, F.,2005. Face Recognition From a Single Image perPerson: A survey, *Pattern Recognition Society*, vol.39, pp.1725-1745.

[30] Wu J. , Zhou Z.H.,2002. Face recognition with one training image per person, Pattern Recognition Lett. 23 (14) 1711–1719.

[31] Yang J., Zhang D., Frangi A.F. , Yang J., 2004. Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (1) 131–137.

[32] Jung H.C., Hwang B.W., Lee S.W., 2004. Authenticating corrupted face image based on noise model, Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 272

[33] Penev, P. S., Atick, J. J., 1996. Local Feature Analysis: A General StatisticalTheory for Object Representation, *Network: Computation in Neural Systems*, 7:3,pp.477-500.

[34] Brunelli R., Poggio T., 1993 . Face recognition: features versus templates, *IEEE Trans. Pattern Anal. Mach. Intell*. Vol.15 (10), pp.1042–1062.

[35] Rowley H.A., Baluja S., Kanade T., 1998. Neural network-based face detection, *IEEE Trans. Pattern Anal. Mach. Intell*. Vol.20 (1), pp.23–38.

[36] Lanitis A., Taylor C.J., Cootes T.F.,1995. Automatic face identification system using flexible appearance models, *Image Vis. Comput*. Vol.13 (5), pp.393–401.

[37] Jain A.K., Zhong Y., sLakshmanan S., 2000. Object matching using deformable templates, *IEEE Trans. Pattern Anal. Mach. Intell.* Vol.22(1), pp.4–37.

[38] Manjunath B.S., Chellappa R., Malsburg C.V.D., 1992. A feature based approach to face recognition, in: Proceedings, *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 373–378.

[39] Lee T.S., 1996. Image representation using 2-d Gabor wavelets*, IEEE Trans. Pattern Anal. Mach. Intell*. Vol.18 (10) , pp.959–971.

[40] Wiskott, L., Fellous, J., Krüger, N., von der Malsburg, C., 1997. Face Recognition by Elastic Bunch Graph Matching, *IEEE Trans. on PAMI*, vol. 19, no. 7, pp.775-779.

[41] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, 1993. Malsburg von der, R. Wurtz, Distortion invariant object recognition in the dynamic link architecture, IEEE Trans. Comput. 42 (3) (300–311.

[42] Nefian, A.V., Hayes III, M. H., 1998. Face Detection and Recognition Using Hidden Markov Models, *In Proceedings of the 1998 IEEE International Conference on Image Processing*, Chicago, Illinois, USA, October 4-7, vol. 1, pp.141-145.

[43] Huang, J., Heisele, B., Blanz, V., 2003. Component-Based Face Recognition with 3D Morphable Models, *In Proc. of International Conf. On Audio- and Video-Based Person Authentication*.

[45] P.N. Belhumeur and D.J. Kriegman,v. What Is the Set of Images of an Object under All Possible Illumination Conditions, Int'l J.Computer Vision, vol. 28, no. 3, pp. 245-260.

[46] H. Wang, S.Z. Li, Y. Wang, and W. Zhang, 2003. Illumination Modeling and Normalization for Face Recognition, Proc. ICCV Workshop on Analysis and Modeling of Faces and Gestures.

[47] Georghiades, A.S., Kriegman, D.J., Belhumeur, P. N. ,2001. "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose", PAMI

[48] Gokmen, M. Jain, A.K. , 1997."λ-τ Space Representation of Images and Generalized Edge Detection," *IEEE Trans. on PAMI*, vol.19, No. 6, pp. 545-563.

[49] Lanitis, A., Taylor, C. J., Cootes, T. F., 2002. Towards Automatic Simulation of Aging Effects on Face Images, *IEEE Trans. on PAMI*, vol. 24, no.4, pp.442-455.

[50] Georghiades, A.S. Belhumeur, P.N. Kriegman, D.J. ,2000, From few to many: generative models for recognition under variable pose and illumination, A*utomatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on***,** pp. 277-284

CURRICULUM VITAE

Candidate's full name	**:**	**Kıvılcım HELHEL**
Place and date of birth	**:**	**ISPARTA 22.04.1983**
Permanent Address	**:**	**İdealtepe Mah. Beyazköşk Sok.**
		No:13/12 Maltepe/İSTANBUL
Universities and Colleges attended	**:**	**Fatih Science School (ESKİŞEHİR) 2002**
		Hacettepe University (ANKARA) 2006